Using a support vector machine and a land surface model to estimate large-scale passive microwave temperatures over snow-covered land in North America

Journal:	*Journal of Selected Topics in Applied Earth Observations and Remote Sensing*
Manuscript ID:	Draft
Manuscript type:	Regular
Date Submitted by the Author:	n/a
Complete List of Authors:	Forman, Barton; University of Maryland, Department of Civil and Environmental Engineering Reichle, Rolf; NASA, GSFC
Keywords:	Modeling, Remote sensing, Snow

Using a support vector machine and a land surface model to estimate large-scale passive microwave temperatures over snow-covered land in North America

Barton A. Forman and Rolf H. Reichle

Abstract

A support vector machine (SVM), a machine learning technique developed from statistical learning theory, is employed for the purpose of estimating passive microwave (PMW) brightness temperatures over snow-covered land in North America as observed by the Advanced Microwave Scanning Radiometer (AMSR-E) satellite sensor. The capability of the trained SVM is compared relative to the artificial neural network (ANN) estimates originally presented in [14]. The results suggest the SVM outperforms the ANN at 10.65 GHz, 18.7 GHz, and 36.5 GHz for both vertically- and horizontally-polarized PMW radiation. When compared against daily AMSR-E measurements *not* used during the training procedure and subsequently averaged across the North American domain over the 9-year study period, the root mean squared error in the SVM output is 8 K or less while the anomaly correlation coefficient is 0.7 or greater. When compared relative to the results from the ANN at any of the six frequency and polarization combinations tested, the root mean squared error was reduced by more than 18% while the anomaly correlation coefficient was increased by more than 52%. Further, the temporal and spatial variability in the modeled brightness temperatures via the SVM more closely agrees with that found in the original AMSR-E measurements. These findings suggest the SVM is a superior alternative to the ANN for eventual use as a measurement operator within a data assimilation framework.

Index Terms

AMSR-E, brightness temperature, modeling, support vector machines, remote sensing, passive microwave, snow

B. A. Forman is with the Department of Civil and Environmental Engineering, University of Maryland, 1173 Glenn Martin Hall, College Park, Maryland, 20742 USA. E-mail: baforman@umd.edu.

R. H. Reichle is with the Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Code 610.1, Greenbelt, Maryland 20771, USA.

I. Introduction and Background

Snow is a critical component of the hydrologic cycle because of its influence on land surface albedo [19], its control on land surface water and energy balances [31], and its impact on weather and climate [4], [17]. Snow also serves as the dominant source of freshwater supply for more than one billion people globally [3], [16]. Direct quantification of the mass of snow, or snow water equivalent (SWE), however, is complicated by significant spatial and temporal variability such that sparse, ground-based observation networks can not always capture the spatiotemporal heterogeneity of SWE. In response, researchers have begun using space-based instrumentation in conjunction with land surface models (LSMs) in an effort to better quantify this vital resource.

Data assimilation can be used to merge satellite-derived measurements with physically-based LSMs [9], [10], [13], [29] by weighing the uncertainties in each in order to yield a merged estimate superior to the measurements or the model alone [25]. In this process, it is necessary to map the relevant model state variables into the corresponding measurement space: In the context of snow data assimilation, this can involve mapping model state variables into passive microwave (PMW) brightness temperature (T_b) space [2], [10], [12] using a physically-based radiative transfer model (RTM) [27], [35], [36]. However, LSMs operating at regional and continental scales do not possess the fidelity to provide the necessary inputs required by the RTM [11], and as such, previous PMW T_b studies have been limited to point-scale or basin-scale applications [2], [10], [12].

Recent research has explored the use of machine learning as an efficient alternative to a RTM in order to map model state variables into PMW T_b space. It was shown that an artificial neural network (ANN) could effectively diagnose PMW T_b at multiple frequencies and multiple polarizations across regional and continental scales [14]. Further, these results were unbiased over the 9-year study period, demonstrated significant skill during both the accumulation (i.e., when the snow is relatively dry) and ablation (i.e., when the snow is relatively wet) phases of the snow season, and yielded a domain-averaged root mean squared error (RMSE) less than 10 K at all frequency and polarization combinations investigated in the study. The findings of [14] were the first to demonstrate the potential of using an ANN as a measurement operator to estimate PMW T_b over snow-covered land with the eventual goal of applying it in a large-scale SWE data assimilation framework.

This current study expands on the work of [14] by investigating an alternative form of machine learning. Namely, the objective of this study is to explore the utilization of a support vector machine (SVM) for nonlinear regression as applied to PMW T_b estimation over snow-covered land, and to contrast the results against those generated by the ANN presented in [14]. SVMs are similar to ANNs in that both forms of machine learning are skilled at reproducing nonlinear processes [8], [26], [39]. However, there are also differences in performance between SVMs and ANNs. For example, if the problem is strictly convex, then the solution to the SVM optimization problem is unique. With convex constrained optimization problems, it has also been shown that SVMs are not plagued with the problem of *local minima* as are ANNs [32]. Further, a number of resampling procedures are available [8] that easily allow for the proper selection of SVM parameters without the need for an "expert" user to decide *a priori* what the SVM parameters should be, which is contrary to the general ANN application case.

The SVM methodology and experimental domain used in this study are outlined in section II and appendix, the approach to validate the results is discussed in section III, the results are presented in section IV, and the major findings and conclusions of this study are highlighted in section V.

II. Methodology

A. SVM Solution

Consider an $[1 \times n]$ input vector, $\mathbf{y}$, where $n = 11$ is the number of geophysical variables that characterize snow and near-surface environmental conditions at a given location in space and time. In this study, $\mathbf{y}$ is derived from a land surface model simulation (further details provided in section II-B). Once trained on T_b observations, a nonlinear SVM can be used to estimate T_b at a given frequency and polarization for a particular location in space and time as a function of $\mathbf{y}$ via the approximating function

$$f(\mathbf{y}) = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i)\, k(\mathbf{x}_i, \mathbf{y}) + \delta \quad (1)$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$ are the $[m \times 1]$ set of dual Lagrangian multipliers, $k(\mathbf{x}_i, \mathbf{y})$ is the radial basis kernel function computed as $k(\mathbf{x}_i, \mathbf{y}) = \exp\{-\gamma \|\mathbf{x}_i - \mathbf{y}\|^2\}$, $\mathbf{x}$ is the $[m \times n]$ training matrix, δ is the "bias" coefficient, and m is the number of training targets. The variables α_i, α_i^*, and δ along with the corresponding set of support vectors are all defined during training, which is discussed in more detail in the appendix. It is worth noting here that $\mathbf{x}$ and $\mathbf{y}$ are computed with the same land surface model, but that the two sets are drawn from different periods of time and can therefore be considered independent. Once the approximating function is specified and the SVM has been trained, equation (1) provides a straightforward and computationally inexpensive method to estimate T_b as a function of time given temporally varying near-surface conditions from the land surface model simulation.

B. SVM Inputs and Outputs

Inputs to the SVM are identical to those used in the ANN study. For brevity, only the essential details are discussed here with the acknowledgement that additional details may be found in [14]. Inputs to the SVM included a number of land surface state estimates derived from the NASA Catchment land surface model (Catchment) [21] and are listed in Table I. State variable estimates from Catchment, in general, are comprised of: 1) snow conditions and 2) near-surface air, soil, and vegetation temperatures. The Catchment model was forced by surface meteorological fields acquired from the Modern Era Retrospective-Analysis for Research and Applications (MERRA) product [30]. Daily-averaged Catchment output was generated on the Equal Area Scalable Earth (EASE) grid at a 25km × 25km horizontal resolution. AMSR-E measurements used as training targets and as independent validation were derived on the same 25-km EASE grid; the AMSR-E T_b measurements are discussed in more detail in section II-C1. The LIBSVM library [6] was employed for all SVM training and estimation activities in this study.

C. SVM Training

1) Training Targets: The SVM was trained using AMSR-E measurements collected at three different frequencies – 10.65 GHz, 18.7 GHz, and 36.5 GHz – at both horizontal and vertical polarization. The resulting combination of the three frequencies and two polarizations yielded a total of six different sets of training targets (or outputs) as listed in Table I. These frequency and polarization combinations were selected due to their sensitivity to snow [5], [16], [19] and because the same combinations were used in [14]. The latter enables a direct comparison between ANN and SVM performance, which is one of the main objectives of this study. Three additional AMSR-E channels – 6.9 GHz, 23.9 GHz, and 89.0 GHz – were available for use but were not employed in the SVM framework. This was done in part to maintain continuity with the ANN study and in part due to physical limitations associated with particular frequencies. For example, the 89.0 GHz channel was avoided due to significant atmospheric effects [7] and limitations associated with precipitating clouds [24]. In addition, even though the 23.9 GHz channel has a penetration depth into the snowpack that lies between that of the 18.7 GHz and 36.5 GHz channels, and could therefore provide additional information about snow conditions, its use was avoided due to significant interactions with atmospheric water vapor. Finally, as was similarly conducted in [14], the 6.9 GHz channel was excluded because its effective field of view (75km $\times$ 43km for the 3 dB footprint; [1]) is much greater than the grid spacing of the 25km $\times$ 25km EASE-grid product and because it is relatively insensitive to terrestrial snow [5]. Additional evidence suggests the 6.9 GHz channel is negatively impacted by radio frequency interference [18], which further motivates its exclusion from the selected training targets.

It has been demonstrated that forest cover attenuates PMW emission from the underlying snowpack while simultaneously adding its own contribution to the radiation as measured by the radiometer [34]. Recent research has further shown that AMSR-E snow retrievals that employ PMW T_bs at 36 GHz are adversely impacted by forest effects and that correction strategies can be applied using radiation transfer theory [22]. In this present study, no such correction strategies have been applied. In other words, the AMSR-E T_b measurements used during training (as well as the T_b estimates generated by the trained SVM) over forested regions contain contributions from both the snow and the vegetative canopy. Vegetation corrections were excluded from this study in order to maintain continuity with the approach outlined in [14]. All AMSR-E T_bs used in this study were obtained from http://nsidc.org/data/nsidc-0301.html and are highlighted in [20].

2) Training Approach: A SVM was generated for each T_b frequency and polarization combination listed in Table I. Each SVM was trained separately and independently at each grid cell on the 25 km EASE grid using the available measurements collected by AMSR-E during the 9-year period from 1 September 2002 to 1 September 2011. This 9-year period encompasses approximately 98% of the available AMSR-E data prior to 4 October 2011 when a problem associated with the rotation of the AMSR-E antenna occurred and regular science data collection ceased. Each SVM was trained for a two week (fortnight) period. This approach was used to address the strong seasonality in snow processes [14].

For a given fortnight in a given year, training activities employed the AMSR-E observations for the given fortnight from the other eight years in the training record. That is, training cycled through the 9-year period withholding each year in turn. Consequently, the AMSR-E measurements for the year that were *not* used for training were later

utilized during validation activities discussed below in section III. An identical procedure to define the training dataset was similarly used in the ANN study as discussed in [14].

Tests were conducted using less than eight years of training data; however, the results from these tests (not shown) suggested that, in general, SVM performance improved as more training data were made available. In addition, in order to enhance continuity from one fortnight to the next, a temporal overlap of two weeks was included at both the beginning and end of each training period. Only measurements collected during the nighttime AMSR-E overpass (roughly between 01:00 to 01:30 hours local time) were used during training in order to minimize wet snow effects.

The SVM training procedure consisted of a two-fold training process (similar to that used during ANN training) in an effort to enhance SVM robustness. The two-fold procedure involved the selection of a subset (approximately 50%) of the 8-year training data with which the SVM was first trained. (Note that the training data discussed here are separate from the independent validation data mentioned above and discussed below in section III.) The trained SVM was then used to reproduce the subset of training data, and the mean square error (MSE) was computed between the SVM estimates and the training data subset. The process employing the first subset of training data was repeated across a range of values for the SVM parameters ε and γ (Appendix), each time computing (and storing) the resulting MSE. This procedure was then repeated using the remaining (i.e., the other 50%) of the training data such that no reuse of training data occurred during the two-fold process. As conducted with the first subset of training data, the SVM was trained across a range of ε and γ values and MSE was computed. The combination of ε and γ values that yielded the closest agreement (in a mean-square sense) across the two training exercises conducted thus far was ultimately selected for use during the final SVM training procedure, which employed the entire (8-year) training data set. This final SVM was then used for the remainder of the comparisons described below. Additional tests ranging from a two-fold process up to a ten-fold process were conducted without any significant improvement found beyond the two-fold process. Therefore, the two-fold procedure was ultimately adopted as it incurred the least amount of computational expense without any sacrifice in SVM performance while also maintaining continuity with the ANN study.

D. Study Domain

The study domain shown in Figure 1 encompasses the North American continent poleward of 32°N and is identical to that used in the ANN study [14]. This region was selected because the domain includes all the major snow classes – tundra, taiga, maritime, prairie, alpine, and ephemeral – as defined in [33]. The 9-year study period (1 September 2002 to 1 September 2011) corresponds to nearly the entire AMSR-E measurement record and is likewise identical to that used in the ANN study [14].

III. Validation Approach

Validation of the SVM-derived estimates involved the use of the original AMSR-E measurements *not* used during SVM training (see section II-C2 for more details). For any year of interest, the validation set of AMSR-E

measurements is completely separate and autonomous from the training datasets, and therefore constitutes a valid, independent comparison. Several different validation metrics were employed: 1) bias of the estimator, *bias*, 2) root mean squared error, $RMSE$, which includes the bias, and 3) anomaly correlation coefficient, anomaly R. The first two metrics were calculated from an original (i.e., "raw") time series. The anomaly R metric, on the other hand, was calculated from an anomaly time series after the respective climatological (multi-year average) seasonal cycle was subtracted from each respective data set. Each metric was computed separately at each grid cell (based on daily data). Area-averaged metrics were computed by averaging the metrics across the snow-covered grid cells.

In order to compute meaningful statistics, a number of constraints were enforced to ensure that time series of sufficient length were available. For example, snow must be present at a given location at least 5% of the year. As a result, the number of data points used in the statistical calculations shown for a grid cell ranged from a minimum of 164 along the southern boundary of the snow covered area to more than 2500 near the northern edge of the study domain. It is well recognized that the AMSR-E measurements contain error (standard deviation of ~1 K according to http://nsidc.org/data/docs/daac/amsre_instrument.gd.html), but this error is small when compared to the uncertainty in the SVM and ANN output (relative to the AMSR-E measurements) and is therefore neglected here. An identical approach was employed in [14] during the original ANN investigation and is similarly applied here to the SVM.

IV. Results

Assessment of SVM capability in estimating AMSR-E T_bs included comparisons of both SVM and ANN output relative to AMSR-E measurements *not* used during training activities. These comparisons included statistical maps for the 9-year study period, which yielded a large-scale analysis of SVM performance relative to the ANN (subsection IV-A). In addition, time series investigations (subsection IV-B) are provided at several different locations (location markers provided on Figure 1) over the course of an entire snow season. The time series investigation provided evidence as to the capability of the machine learning techniques at reproducing AMSR-E measurements during both the snow accumulation portion of a snow season and the subsequent ablation phase. Moreover, a brief investigation on the spatial and temporal variability of the machine learning estimates is provided in section IV-C in order to highlight each technique's skill at reproducing the variability in the original AMSR-E measurements. Finally, the potential for employing the SVM within a data assimilation framework (subsection IV-D) is briefly highlighted via investigation of the resulting Kalman gain matrix. In an analogous manner as conducted in [14], most discussions focused on the 18V and 36V results because these channels are considered the most informative when viewed in the context of SWE estimation [5]. However, it is worth noting here that all frequency and polarization combinations listed in Table I were investigated and analyzed in a similar fashion as the 18V and 36V results.

A. Cross-validation

Figure 2 provides a large-scale overview of SVM versus ANN performance at 18V over the course of the 9-year study period. Each subplot represents a statistical map for either the SVM output (left column) or the ANN

output (middle column) computed relative to the AMSR-E measurements *not* used during training. The top row illustrates the *bias* in the SVM estimates (Figure 2a), the *bias* in the ANN estimates (Figure 2b), and the difference between the two (Figure 2c). Analogously, the middle row highlights the computed $RMSE$ whereas the bottom row highlights the computed anomaly R.

In terms of *bias*, both the SVM and ANN yield relatively unbiased estimates when averaged over the entire study domain across the 9-year study period. The SVM estimates contain approximately 1 K more positive *bias* (relative to the ANN estimates) in regions surrounding Hudson Bay, across northern Quebec, and in western Alaska near the Bearing Sea. Conversely, the SVM contains approximately 1-2 K more negative *bias* in regions covered by boreal forest. Figure 2c further highlights the increase in the magnitude of *bias* in the SVM output (relative to the ANN estimates), but this bias is small when compared to the temporal variability of the original AMSR-E measurements (further discussion provided in section IV-C) and, in general, falls within the estimated error standard deviation (~1 K according to http://nsidc.org/data/docs/daac/amsre_instrument.gd.html).

Despite the small increase in *bias* generated by the SVM relative to the ANN output, results provided in Figures 2d-f show the SVM contains significantly less $RMSE$ than the ANN. The reduction in $RMSE$ within the SVM estimates is witnessed across the entire study domain, including regions with and without significant vegetative cover, and are most apparent in regions where sub-grid scale lakes (i.e., lakes smaller than the 25-km EASE pixel size) are common. Additional reductions in $RMSE$ are also found along the southern periphery of the snow line where the snow pack is thin and ephemeral and where freeze-thaw cycles are relatively common [14].

Figure 3a presents box plots of computed $RMSE$ for all frequency and polarization combinations examined in this study. It is clear that the SVM yields a reduction in computed $RMSE$ relative to the ANN results. When viewed from the perspective of the median value, SVM-derived $RMSE$ is reduced, on average, by ~20-25% from the ANN-derived results. In addition, the extreme values (i.e., the 90th-percentiles) are greatly reduced such that the SVM yields more stable and more accurate results when compared to the ANN estimates for the same study period and study domain.

The final set of statistics provided in Figures 2g-i shows the computed anomaly R over the 9-year study period. Anomaly R is useful in that it focuses on the capability of each technique to capture the synoptic-scale and inter-annual variability of the T_b estimates across the entire spatial domain. As is clearly seen, the SVM-based estimates are superior to those derived from the ANN. In particular, the anomaly R in regions to the north and south of the boreal forest is nearly doubled from ~0.4 to ~0.8. Within forested regions, T_b as measured by AMSR-E includes PMW emission from the forest canopy [22]. The ANN benefits greatly from model-derived skin temperature, T_{skin}, within the forest canopy, which yields much greater anomaly R values in regions where significant forest cover is present [14]. However, in regions where significant forest cover is not present, ANN-based performance as a function of time is drastically reduced. The SVM, on the other hand, is able to better utilize the full set of input variables outlined in Table I across a broader range of conditions, including both forested and non-forested areas. The dramatic improvements in anomaly R values computed from the SVM for the other evaluated frequency and polarization combinations are further witnessed in Figure 3b. The SVM is clearly able to capture much more

of the temporal variability found in the original AMSR-E T_b measurements. As was the case with the $RMSE$ results, the median anomaly R values based on the SVM estimates are better for the vertically-polarized channels when compared to the horizontally-polarized channels, but these differences are relatively small and suggest that, in general, the SVM outperforms the ANN across space and time at all frequency and polarization combinations evaluated in this study.

B. Time Series Investigation

Results presented thus far focused on the time-integrated behavior of the SVM over the 9-year study period and its performance relative to that of the ANN. A time series investigation is discussed here in order to better illustrate the performance of the SVM throughout the snow season at a handful of representative locations. The goal of this investigation is to highlight the capability of the SVM to estimate T_b during the snow accumulation season when the snow is dry (and hence acts as an efficient scatterer) as well as during the snow ablation season when the snow is relatively wet (and hence acts a relatively efficient emitter). The 2003 – 2004 snow season was selected for analysis because it is representative of a typical snow season during the 9-year study period.

Figure 4 highlights T_b time series for three different locations (shown as red circles in Figure 1). These particular locations were chosen because they represent the most dominant snow classifications (in terms of North American coverage in Figure 1) and because these three locations represent a range of different vegetative covers as well as maximum snow depths at peak accumulation. Namely, the first subplot (Figure 4a) is for a location with relatively shallow snow and little vegetative cover, the second subplot (Figure 4b) is for a location with moderate snow depth and relatively thick vegetative cover, and the third subplot (Figure 4b) is for a location with relatively deep snow and a modest amount of vegetative cover. The short gap in all time series in early-November 2003 is due to missing AMSR-E observations. The presence of a solid line (ANN or SVM) indicates the presence of snow as modeled by Catchment.

As is shown, both the ANN and the SVM do a reasonable job at reproducing the AMSR-E measurements *not* used during training. Both techniques of machine learning capture the large-scale features present in the AMSR-E T_b measurements, including both the accumulation and ablation phases of the snow season. However, clear differences between the ANN and SVM estimates are also seen. Namely, the SVM does a much better job of capturing the high frequency (i.e., day-to-day) variability associated with synoptic scale processes. The ANN estimates, on the other hand, often lack this high frequency variability as is witnessed by the step function-like features present during portions of the snow season at each of the three locations. These clear differences in the ANN versus SVM estimated variability over time scales of a few days to a week corroborate the anomaly R results highlighted in Figures 2g-i and Figure 3b. These findings suggest the ANN output is less sensitive to certain changes in the modeled inputs whereas the SVM output is significantly more sensitive to changes in the modeled inputs as a function of time, and hence, yield T_b estimates that capture more of the high frequency temporal variability. Similar features are also found in the 10H, 10V, 18H, and 36H T_b estimates (results not shown).

An additional note of interest regards the presence of snow as modeled by Catchment, which is used as input to

both the ANN and SVM (Table I). The AMSR-E measurements shown in Figures 4a and 4b suggest the presence of snow when the difference between 18V and 36V is greater than zero. The Catchment model, in general, suggests the presence of snow only where ANN or SVM predictions are made available (i.e., by the solid lines). In Figures 4a and 4b, the Catchment model predicts the complete melt of the snow pack several weeks earlier than is suggested by the AMSR-E measurements. The exact cause of the discrepancy is currently unknown (and beyond the current scope of work for this study). It remains to be seen whether such errors could be corrected through data assimilation.

C. Output Variability

The findings presented above demonstrate the ability of a SVM to yield relatively unbiased AMSR-E T_b estimates with a modest amount of $RMSE$ and significant skill (in terms of anomaly R) over synoptic and seasonal time scales. Further, it was shown the SVM improves upon the ANN, in general, at all frequency and polarization combinations examined in this study. An important question that remains, especially when viewed in the context of a data assimilation framework, is whether the SVM estimates can reasonably represent the spatiotemporal variability of the AMSR-E T_b measurements.

The bar plots in Figure 5 highlight the variability for both the ANN-derived and SVM-derived T_b estimates. The corresponding variability in the AMSR-E measurements *not* used during training is also included. In addition to showing results for all of the frequency and polarizations used in this study, the results are further stratified by snow class (Figure 1). Each of the six snow classes – tundra, taiga, maritime (abbreviated mari.), alpine, prairie, and ephemeral (abbreviated ephem.) – cover hundreds (or more) EASE grid cells. For the purpose of this analysis, variability is first computed as the spatial standard deviation for each day when snow is present and then averaged in time over the 9-year study period. As is shown in Figure 5, both the ANN and SVM variabilities agree quite well with the variability in the AMSR-E measurements for each frequency/polarization combination and for each snow class. However, it is clear the SVM agrees better with the AMSR-E measurements (relative to the ANN-based output) in almost every category. In addition, the SVM (and ANN) estimates capture many of the large-scale features witnessed in the AMSR-E measurements. For example, the variability in the horizontally-polarized estimates is generally greater than their vertically-polarized counterparts for a given snow class. This behavior can be partly explained by the increased sensitivity of horizontally polarized T_bs to the presence of internal ice layers and surface crust [24], [28]. Further, the variability in both the ANN and SVM estimates (and AMSR-E measurements) is generally greatest in the tundra and taiga regions where the boreal forest is located, which suggests the forest influences contained within the AMSR-E measurements [22] are reproduced by the machine learning techniques. However, the SVM estimates clearly match the AMSR-E measurements more closely (relative to the ANN) for all frequency, polarization, and snow class combinations. The increased variability in the SVM estimates (relative to the ANN) corroborates the previous results that showed the SVM captures much more of the high frequency variability at a given location (see Figure 4 for examples), which leads to the increased variability across space and time as witnessed in Figure 5.

D. Potential for Data Assimilation

The development of the SVM was originally motivated so that it could eventually be included as an observation operator within a data assimilation framework [25], [29] for the purpose of merging AMSR-E T_b measurements with SVM-derived T_b estimates into a LSM. In order to assess the potential of the SVM within the data assimilation framework, a brief investigation of the error covariance structure between the LSM and the SVM-based T_b estimates is presented here. The error covariance is computed as a gain matrix, **K**, which represents a weighted average of the uncertainty in the LSM-derived estimates of SWE along with the spectral difference in the T_b estimates at 18V and 36V. The presence of a non-zero error covariance structure would suggest a degree of potential for a follow-on data assimilation study employing the SVM. The gain **K** is computed as

$$\mathbf{K} = \mathbf{C}_{yz} \left(\mathbf{C}_{zz} + \mathbf{C}_{vv}\right)^{-1}, \tag{2}$$

where $\mathbf{C}_{yz}$ is the (sample) cross-covariance between the ensemble of prior land model states and the SVM- or ANN-predicted T_bs, $\mathbf{C}_{zz}$ is the (sample) covariance of the predicted T_bs, and $\mathbf{C}_{vv}$ = 1 K^2 is the T_b measurement error variance. The gain K is computed at each pixel between the modeled SWE and SVM-based estimates of ΔT_b=18V-36V. The spectral difference ΔT_b=18V-36V is employed here as it is commonly used to estimate SWE [15] and serves to represent the linkage between SWE and T_b. The larger the spectral difference, in general, the greater amount of SWE is present [5]. Catchment model perturbations were implemented using the methods of [13] and performed in an identical fashion as conducted in [14]. As a first approximation for demonstrating the *potential* for a non-zero error covariance structure, only the prior model estimate (without an analysis update step) is used here. Again, this simplified approach is merely to demonstrate the *potential* for future inclusion into a data assimilation procedure.

Figure 6 shows the computed gain over snow-covered regions in the domain collocated with the presence of AMSR-E measurements on 1 February 2003 when SWE is near peak accumulation. The collocation with AMSR-E serves to highlight the spatial extent that could eventually be updated when using the Kalman filter. Figure 6a shows the gain using the ANN-based T_b estimates whereas Figure 6b show the computed gain using the SVM-based T_b estimates. If the difference between the AMSR-E measurements and the T_b estimates is +1 K, a gain of **K**=10 mm K^{-1} translates to an increase of 10 mm in the posterior (updated) modeled SWE. Alternatively, if the gain is negative (e.g., **K**=-10 mm K^{-1}) with a +1 K difference between the AMSR-E measurement and the estimated T_b would result in a decrease in the posterior SWE estimate of 10 mm.

The large-scale structure in Figures 6a and 6b are similar in that relatively large, positive gains are found in the northeastern portion of the domain as well as throughout much of the Rocky Mountains. In both cases, this is due to relatively small values of $\mathbf{C}_{zz} + \mathbf{C}_{vv}$ in conjunction with relatively large values of C_{yz}. However, significant differences also occur as quantified by a modest pattern (spatial) correlation of R = 0.31 between the maps shown in Figure 6. More specifically, the ANN-derived gain in Figure 6a contains a series of unusual striations across the north-central portion of Canada. These striations are apparently the result of limited sensitivity in ANN output due to small perturbations in an ensemble of ANN inputs. The result is that neighboring cells, at times, yield similar

(or identical) values of **K**, which can, at times, result in the appearance of striated features. The SVM-based gain in Figure 6b, on the other hand, does not suffer from these striated features and, in turn, yields a more smooth and continuous estimate of the computed gain across space. The presence of negative gains is a particularly interesting feature given the first SWE retrieval algorithm originally presented by [5]. Namely, the earliest retrieval algorithm suggested a direct, linear relationship between SWE and ΔT_b. However, the presence of positive and negative gains shown here suggests a non-linear relationship between SWE and ΔT_b. More work is required to better understand this non-linear behavior, but is considered beyond the scope of the present study. Even though this simple exercise is far from the in-depth investigation of error covariance planned for a follow-on study, it does serve to demonstrate that a non-zero error covariance exists between the modeled SWE and the estimated ΔT_b spectral difference and that this error structure could be leveraged within an ensemble filter framework in order to produce a merged (updated) model estimate of SWE that improves upon the original (prior) model estimate.

V. Conclusions

An SVM was developed in order to estimate AMSR-E T_b at specific frequencies and polarizations. The eventual use of the SVM is to serve as a measurement operator within an ensemble-based data assimilation framework for the purpose of improving SWE estimation at regional and continental scales. The model capability of the SVM was compared against an alternative form of machine learning – the ANN – originally presented in [14]. Quantitative comparisons are made to highlight the skill of the SVM relative to that of the ANN. Both the SVM and ANN utilize output from the NASA Catchment model (forced with MERRA surface meteorological fields) as input for generating T_b estimates. Horizontally- and vertically-polarized T_bs from AMSR-E at 10.65, 18.7, and 36.5 GHz supplied on a 25km $\times$ 25km resolution equal area grid were used during training. Subsequent comparisons with SVM and ANN estimates employed AMSR-E measurements *not* used during the training activities so that independent verification activities could be conducted.

When averaged across the North American study domain over the course of a 9-year study period, SVM-derived T_b estimates were found to be relatively unbiased ($|bias| \lesssim 1$ K), contain median $RMSE$ values of less than 10 K, and possess skill that yielded anomaly R values on the order of 0.8. The SVM technique outperformed the ANN in every major snow class (as defined by [33]) with a notable increase in the ability to reproduce the high frequency temporal variability present in the AMSR-E measurements. In addition, a brief inspection was made into the error covariance structure between modeled SWE (via the Catchment model) and a spectral difference in T_b as computed by the two different machine learning techniques. The results showed the presence of a non-zero covariance structure, which could eventually be leveraged within a data assimilation framework in order to improve regional- and continental-scale SWE estimates.

In short, the trained SVM presented here is a superior alternative to the ANN originally presented in [14]. Even though the training data used by both techniques were identical, and all other relevant aspects during the learning process were held as equivalent as possible between the two different machine learning techniques, it is clear that the SVM as applied in this study yields better performance. One hypothesis is that the SVM learning procedure

focuses on a single frequency / polarization combination whereas the ANN simultaneously yields estimates for all of the frequency / polarization combinations. The reduction in the number of degrees of freedom in the ANN is a likely contributor to the reduced performance relative to the SVM. A series of tests (results not shown) using an ANN trained to estimate only a single frequency / polarization combination found improved performance relative to the multi-T_b presented in [14]. However, the level of improvement using the single T_b ANN framework still did not achieve the same degree of performance as found with the SVM. The increase in degrees of freedom in the SVM relative to the ANN, in part, helps explain why the SVM outperformed the ANN. An additional reason for this behavior could also be attributed the dependence of input from an "expert" user regarding the exact structure of the ANN (e.g., number of layers, number of hidden nodes per layer) prior to training that is not similarly required in the SVM setup.

In an analogous manner as the ANN presented in [14], it is worthwhile to discuss and highlight some of the limitations of the SVM presented here. For starters, the Catchment model used to generate the inputs to the SVM does not account for ice crust on the surface of the snowpack, internal ice layers within the snowpack, or sub-grid scale lake ice underlying the snowpack. Hence, the SVM-derived estimates do not explicitly account for these effects, which limits the skill of the SVM-based T_b estimates. In addition, AMSR-E is no longer collecting measurements due to a problem associated with the rotation of the AMSR-E antenna. However, AMSR2 on-board the Japanese Global Change Observation Mission – Water (GCOM-W) satellite is currently collecting T_b measurements at comparable frequencies and polarizations as AMSR-E before its malfunction. Preliminary results to be presented in a follow-on study suggest the SVM (and ANN) can be trained on AMSR-E measurements and then used to subsequently predict AMSR2 T_bs. That is, the machine learning technique can be used to estimate measurements from one sensor using training targets from another sensor. This transferability could enable a continuous record forward in time even though AMSR-E science data collection is inactive.

Despite some deficiencies in the SVM approach, it is worthwhile reiterating the skill in the SVM estimates and the clear improvements relative to the ANN-based approach in [14]. The SVM was shown to effectively reproduce AMSR-E T_bs at multiple frequencies and polarizations during both the accumulation phase when the snowpack relatively dry as well as during the ablation phase when the snowpack is relatively wet. Significant skill was demonstrated in both shallow and deep snow environments, in areas with and without vegetative cover, and across all six major snow classes common across North America (and the northern hemisphere as a whole). On-going studies are investigating the sensitivity of SVM-derived estimates to snow-related variables (most notably SWE) and preliminary results suggest a considerable amount of sensitivity is present in the SVM across the majority of the study domain. These findings suggest a trained SVM could serve as an effective and computationally efficient measurement operator within a data assimilation procedure for which it was originally constructed.

Acknowledgments

The AMSR-E brightness temperature measurements were obtained from the National Snow and Ice Data Center Distributed Active Archive Center (NSIDC DAAC), University of Colorado at Boulder. The LIBSVM library was

provided by National Taiwan University. Partial financial support for B. A. Forman was provided by the Minta Martin Foundation. R. H. Reichle was supported by the NASA program on the Science of Terra and Aqua.

Appendix

Consider an $[m \times n]$ training matrix, $\mathbf{x}$, and an $[m \times 1]$ vector of training targets, $\mathbf{z}$, such that $\{(\mathbf{x}_1, z_1), \ldots, (\mathbf{x}_m, z_m)\}$. In the context of this study, $\mathbf{x}$ represents $n = 11$ geophysical variables that characterize snow and near-surface environmental conditions at a given location and at m different times as derived from a land surface model simulation (further details provided in section II-B). The vector $\mathbf{z}$ represents a corresponding series of m satellite-based measurements of PMW T_b at a given frequency and polarization (further details provided in section II-C1). Assume $\phi(\mathbf{x})$ is a nonlinear function that maps the geophysical inputs from the land surface model, $\mathbf{x}$, into T_b space as

$$f(\mathbf{w}, \delta) \quad = \quad \langle \mathbf{w} \cdot \phi(\mathbf{x}) \rangle + \delta \tag{3}$$

where $\mathbf{w}$ is a vector of weights, $\langle \mathbf{w} \cdot \phi(\mathbf{x}) \rangle$ is the inner dot product of $\mathbf{w}$ and $\phi(\mathbf{x})$, and δ is a "bias" coefficient. For given parameters $C > 0$ and $\varepsilon > 0$, the standard (primal) form of nonlinear support vector regression [6], [37] may be written as

$$\begin{aligned} \underset{\mathbf{w},\, \delta,\, \boldsymbol{\xi},\, \boldsymbol{\xi}^*}{\text{minimize}} \quad & \frac{1}{2}\langle \mathbf{w} \cdot \mathbf{w} \rangle + C \sum_{i=1}^{m} (\xi_i + \xi_i^*) \\ \text{subject to} \quad & \langle \mathbf{w} \cdot \phi(\mathbf{x}_i) \rangle + \delta - z_i \leq \varepsilon + \xi_i, \\ & z_i - \langle \mathbf{w} \cdot \phi(\mathbf{x}_i) \rangle - \delta \leq \varepsilon + \xi_i^*, \\ & \xi_i, \xi_i^* \geq 0, i = 1, 2, \ldots, m. \end{aligned} \tag{4}$$

where m is the available number of T_b measurements in time (for a given location in space), z_i is a T_b measurement at time i, and $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^*$ are slack variables. The values of $\mathbf{w}$, δ, $\boldsymbol{\xi}$, and $\boldsymbol{\xi}^*$ are not specified *a priori*, but rather are determined as a result of the minimization process. The goal of the minimization procedure is to determine values for $\mathbf{w}$, δ, $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^*$ such that the mapped inputs (computed as $\langle \mathbf{w} \cdot \phi(\mathbf{x}_i) \rangle + \delta$) most closely agree with the training targets, $\mathbf{z}$, provided in T_b space.

The primal optimization is commonly solved in dual form [6], [32], [37], [40] by differentiating the primal form with respect to the primal variables (i.e., $\mathbf{w}$, δ, $\boldsymbol{\xi}$, and $\boldsymbol{\xi}^*$) as follows:

$$\begin{aligned} \underset{\alpha_i,\, \alpha_i^*}{\text{minimize}} \quad & \frac{1}{2} \sum_{i,j=1}^{m} (\alpha_i - \alpha_i^*)\left(\alpha_j - \alpha_j^*\right) \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle \\ & + \varepsilon \sum_{i=1}^{m} (\alpha_i + \alpha_i^*) - \sum_{i=1}^{m} z_i (\alpha_i - \alpha_i^*) \\ \text{subject to} \quad & \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) = 0, \\ & \alpha_i \,,\, \alpha_i^* \in [0 \,,\, C] \,,\; i = 1, 2, \ldots, m \end{aligned} \tag{5}$$

where α_i and α_i^* are a dual set of Lagrangian multipliers, $\langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle$ is the inner dot product of $\phi(\mathbf{x}_i)$ and $\phi(\mathbf{x}_j)$, ε is the specified error tolerance, and C is a positive constant that dictates the penalized loss during SVM training. The Lagrangian multipliers, α_i and α_i^*, are nonzero for points equal to or outside of the ε-insensitive tube and alternatively vanish for points inside the ε-insensitive tube. The points with nonzero Lagrangian multipliers comprise the so-called "support vectors". The process described here is similar to that employed by an ANN with a fundamental difference in that the SVM utilizes the weights (computed as $\alpha_i - \alpha_i^*$) as a subset of the training patterns [32].

The computation of $\langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle$ in feature space is often too complex to perform [32]. However, the computation may be conducted in input (land surface model) space using the kernel function $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle$ in order to yield the inner products in feature space, which helps avoid problems of computational infeasibility associated with directly evaluating the basis function in high dimensionality feature space. In this particular study, a radial basis kernel function, $k(\mathbf{x}_i, \mathbf{x}_j)$, was employed that satisfies the expression $k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j) \rangle = \exp\{-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\}$ where $\mathbf{x}_i$ and $\mathbf{x}_j$ are single instances of $\mathbf{x}$ (in time and space), $\|\cdot\|$ represents the Euclidean norm, and γ is proportional to the inverse square of the width parameter as described in [8]. The loss function was specified as ε-insensitive [37], [38]. Quadratic, Laplace, and Huber loss functions were also tested. Since no notable improvements over the ε-insensitive loss function were found, the ε-insensitive loss function was selected as the most appropriate. In addition, the regularization parameter, C, was defined as the range of the training targets (i.e., $C = \max\{\mathbf{z}\} - \min\{\mathbf{z}\}$) using the methods outlined in [23]. An alternate formulation based on [8] was tested using $C = 6\sigma_{\mathbf{z}}$, where $\sigma_{\mathbf{z}}$ is the standard deviation of the training targets, but no significant difference between the two different definitions of C was found. Hence, the former approach was employed such that C was set equal to the range of the training targets. Once the solutions for α_i and α_i^* are found, estimates of T_b can then be computed via Equation (1) using geophysical inputs (derived from the land surface model), $\mathbf{y}$, that are distinct from the training data where $\mathbf{x}$ represents the training matrix and δ is computed as the average of the support vectors (i.e., the subset of the training data with nonzero Lagrangian multipliers).

References

[1] P. Ashcroft and F. Wentz. Algorithm Theoretical Basis Document (ATBD) AMSR Level 2A Algorithm. RSS Technical Report 121599B-1, pp. 27, 3 Nov 2000.

[2] K.M. Andreadis, D. Liang, L. Tsang, D. P. Lettenmaier, and E. G. Josberger. Characterization of errors in a coupled snow hydrologymicrowave emission model. *J. Hydrometeor.*, 9(1), 149–164, doi:10.1175/2007JHM885.1, 2008.

[3] T.P. Barnett, J.C. Adam, and D.P. Lettenmaier. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature*, 438:303–309, 2005.

[4] T.P. Barnett, L. Dümenil, U. Schlese, E. Roeckner, and M. Latif. The effect of Eurasian snow cover on regional and global climate variations. *J. Atmos. Sci.*, 46(5), 661–686, 1989.

[5] A.T.C. Chang, J.L. Foster, and D.K. Hall. NIMBUS-7 SMMR derived global snow cover parameters. *Ann. Glaciol.*, 9:39–44, 1987.

[6] C.C. Chang and C.J. Lin. LIBSVM: A library for support vector machines. *ACM T. Intell. Sys. Tech.*, 2(3), 27, doi:10.1145/1961189.1961199, 2011.

[7] A.T.C. Chang, and L. Tsang. A neural network approach to inversion of snow water equivalent from passive microwave measurements. *Nord. Hydrol.*, 23:173–182, 1992.

[8] V. Cherkassky and Y. Ma. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks*, 17(1), 113126, doi:10.1016/S0893-6080(03)00169-2, 2004.

[9] G.J.M. De Lannoy, R.H. Reichle, P.R. Houser, K. Arsenault, N.E.C. Verhoest, and V.R.N. Pauwels. Satellite-scale snow water equivalent assimilation into a high-resolution land surface model. *J. Hydrometeorol.*, 11:352–369, 2010.

[10] M.T. Durand and S.A. Margulis. Correcting first-order errors in snow water equivalent estimates using a multifrequency, multiscale radiometric data assimilation scheme. *J. Geophys. Res.*, 112:doi:10.1029/2006JD008067, 2007.

[11] Durand, M., E. J. Kim, and S. A. Margulis. Quantifying uncertainty in modeling snow microwave radiance for a mountain snowpack at the point-scale, including stratigraphic effects. *IEEE Geosci. Remote S.*, 46(6), 1753-1767, doi:10.1109/TGRS.2008.916221, 2008.

[12] M.T. Durand, E. J. Kim, and S.A. Margulis. Radiance assimilation shows promise for snowpack characterization. *Geophys. Res. Lett.*, 36(2),doi:10.1029/2008GL035214, 2009.

[13] B.A. Forman, R.H. Reichle, and M. Rodell. Assimilation of terrestrial water storage from GRACE in a snow-dominated basin. *Water Resour. Res.*, 48(1), doi:10.1029/2011WR011239, 2012.

[14] B.A. Forman, C. Derksen, and R.H. Reichle. Estimating passive microwave brightness temperature over snow-covered land in North America using a land surface model and an artificial neural network. *IEEE T. Geosci. Remote*, doi:10.1109/TGRS.2013.2237913, 2013.

[15] J.L. Foster, C. Sun, J.P. Walker, R. Kelly, A.T.C. Chang, J. Dong, and H. Powell. Quantifying the uncertainty in passive microwave snow water equivalent observations. *Remote Sens. Environ.*, 94:187–203, 2005.

[16] J.L. Foster, D.K. Hall, J.B. Eylander, G.A. Riggs, S.V. Nghiem, M. Tedesco, E. Kim, P.M. Montesano, R.E.J. Kelly, K.A. Casey, and B. Choudhury. A blended global snow product using visible, passive microwave, and scatterometer satellite data. *Int. J. Remote Sens.*, 32:1371–1395, 2011.

[17] G. Gong, D. Entekhabi, J. Cohen, and D. Robinson. Sensitivity of atmospheric response to modeled snow anomaly characteristics. *J. Geophys. Res.*, 109(D6), D06107, doi:10.1029/2003JD004160, 2004.

[18] K. Imaoka, M. Kachi, M. Kasahara, N. Ito, K. Nakagawa, and T. Oki Instrument performance and calibration of AMSR-E and AMSR2. *Int. Arch. Photogramm. Remote Sens. Spatial Info. Sci.*, 38(8), 2010.

[19] R. Kelly. The AMSR-E snow depth algorithm: Description and initial results. *J. Remote Sens. Soc. Japan*, 29(1), 307–317, 2009.

[20] K.W. Knowles, M.H. Savoie, R.L. Armstrong, and M.J. Brodzik. AMSR-E/Aqua daily EASE-grid brightness temperatures, September 2002 through August 2011. Boulder, Colorado USA: National Snow and Ice Data Center. Digital media. 2006, updated 2011.

[21] R.D. Koster, M.J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar. A catchment-based approach to modeling land surface processes in a general circulation model 1. Model structure. *J. Geophys. Res.*, 105:24809–24822, 2000.

[22] A. Langlois, A. Royer, F. Dupont, A. Roy, K. Goïta, and G. Picard. Improved corrections of forest effects on passive microwave satellite remote sensing of snow over boreal and subarctic regions. *IEEE T. Geosci. Remote*, 49:3824–3837, 2011.

[23] D. Mattera, and S. Haykin. Support vector machines for dynamic reconstruction of a chaotic system, in *Advances in Kernel Methods*. Edited by B. Schölkopf, C. J. C. Burges, and A. J. Smola. pp. 211–241, MIT Press, Cambridge, MA, USA, 1999.

[24] C. Mätzler. Passive microwave signatures of landscapes in winter. *Meteorol. Atmos. Phys.*, 54:241–260, 1994.

[25] D.B. McLaughlin. An integrated approach to hydrologic data assimilation: Interpolation, smoothing, and filtering. *Adv. Water Resour.*, 12:1275–1286, 2002.

[26] I.S. Msiza, F.V. Nelwamondo, and T. Marwala. Water demand prediction using artificial neural networks and support vector regression. *J. Computers*, 3(11), 1–8, 2008.

[27] J. Pulliainen. Mapping of snow water equivalent and snow depth in boreal and sub-arctic zones by assimilating space-borne microwave radiometer data and ground-based observations. *Remote Sens. Environ.*, 101:257–269, 2006.

[28] A. Rees, J. Lemmetyinen, C. Derksen, J. Pulliainen, and M. English. Observed and modelled effects of ice lens formation on passive microwave brightness temperatures over snow covered tundra. *Remote Sens. Environ.*, 114:116–126, 2010.

[29] R.H. Reichle, D.B. McLaughlin, and D. Entekhabi. Hydrologic Data Assimilation with the Ensemble Kalman Filter. *Mon. Weather Rev.*, 130:103–114, 2002.

[30] M.M. Rienecker, M.J. Suarez, R. Gelaro, R. Todling, J. Bacmeister, E. Liu, M.G. Bosilovich, S.D. Schubert, L. Takacs, G.-K. Kim, S. Bloom, J. Chen, D. Collins, A. Conaty, A. da Silva, W. Gu, J. Joiner, R.D. Koster, R. Lucchesi, A. Molod, T. Owens, S. Pawson, P. Pegion, C.R. Redder, R.H. Reichle, F.R. Robertson, A.G. Ruddick, M. Sienkiewicz, J. Woolen. MERRA - NASA's modern-era retrospective analysis for research and applications. *J. Clim.*, 24:3624–3648, 2011.

[31] D. A. Robinson, K.F. Dewey, and R.R. Heim. Global snow cover monitoring: An update. *B. Am. Meteorol. Soc.*, 74(9), 1689–1696, 1993.

[32] A.J. Smola and B. Schölkopf. A tutorial on support vector regression. *Stat. Comput.*, 14(3), 199–222, 2004.

[33] M. Sturm, J. Holmgren, and G.E. Liston. A seasonal snow cover classification system for local to global applications. *J. Climate*, 8:1261–1283, 1995.

[34] M. Tedesco, and J.R. Wang. Assessment of the NASA AMSR-E SWE product. *IEEE J. Sel. Top. Appl.*, 3:141–159, 2010.

[35] M. Tedesco, and P.S. Narvekar. Atmospheric correction of AMSR-E brightness temperatures for dry snow cover mapping. *IEEE Geosci. Remote S.*, 3(3), 320–324, 2006.

[36] L. Tsang, Z. Chen, S. Oh, R.J. Marks, and A.T.C. Chang. Inversion of snow parameters from passive microwave remote sensing measurements by a neural network trained with a multiple scattering model. *IEEE T. Geosci. Remote*, 30:1015–1024, 1992.

[37] V.N. Vapnik. *Statisitical Learning Theory*, John Wiley and Sons, New York, pp. 736, 1998.

[38] V.N. Vapnik. *The Nature of Statistical Learning Theory*, 2nd ed., Springer-Verlag Inc., New York, pp. 314, 2000.

[39] W.-C. Wang, K.-W. Chau, C.-T. Cheng, and L. Qiu. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *J. Hydrol.*, 374(3-4), 294–306, doi:10.1016/j.jhydrol.2009.06.019, 2009.

[40] P.-S. Yu, S. T. Chen, and I. F. Chang. Support vector regression for real-time flood stage forecasting. *J. Hydrol.*, 328(3-4), 704716, doi:10.1016/j.jhydrol.2006.01.021, 2006.

TABLE I

SVM INPUTS AND OUTPUTS (REPRODUCED FROM [14]).

Inputs	**Symbol**
Top layer snow density	ρ_{sn1}
Middle layer snow density	ρ_{sn2}
Bottom layer snow density	ρ_{sn3}
Snow liquid water content[a]	$SLWC$
Snow water equivalent[a]	SWE
Near-surface air temperature	T_{air}
Near-surface soil temperature	T_{p1}
Skin temperature	T_{skin}
Top layer snow temperature	T_{sn1}
Bottom layer snow temperature	T_{sn3}
Temperature gradient index	TGI
Outputs	**Symbol**
T_b at 10.65 GHz, H-polarization	$10H$
T_b at 10.65 GHz, V-polarization	$10V$
T_b at 18.7 GHz, H-polarization	$18H$
T_b at 18.7 GHz, V-polarization	$18V$
T_b at 36.5 GHz, H-polarization	$36H$
T_b at 36.5 GHz, V-polarization	$36V$

[a] = Column-integrated quantity;

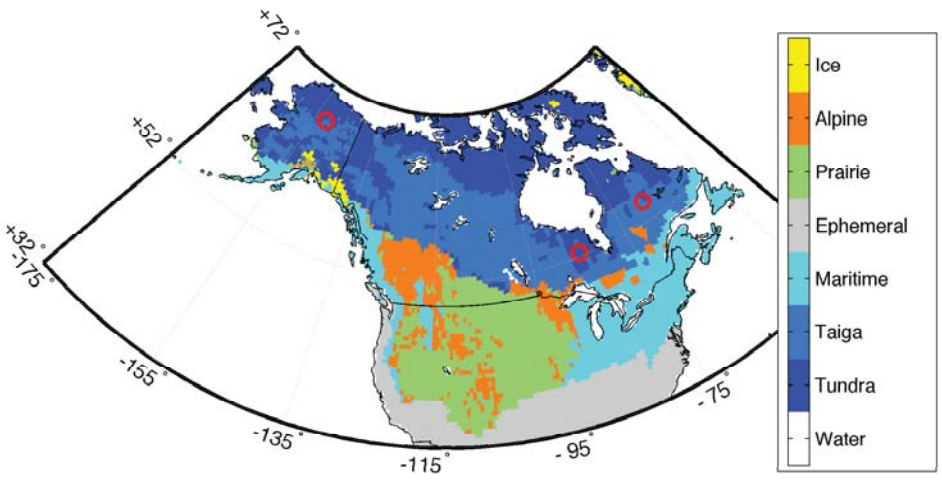

Fig. 1. Study domain encompassing North America poleward of 32° N. Coloring represent the snow classification of [33]. The three red circles represent the locations of the time series comparisons shown in Figure 4.

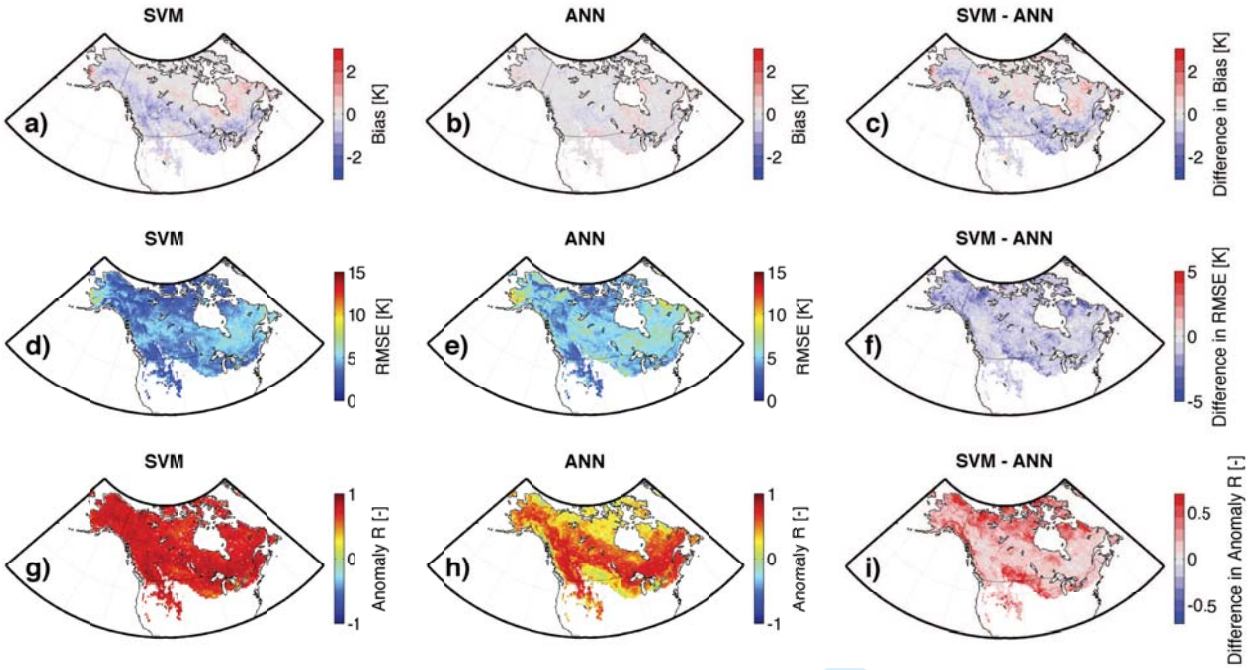

Fig. 2. (Top row) *bias*, (middle row) *RMSE*, and (bottom row) anomaly *R* for the ANN and SVM (versus AMSR-E observations *not* used during training) at 18V for the time period 1 September 2002 through 1 September 2011. Results include (left column) SVM metrics, (middle column) ANN metrics, and (right column) computed difference between SVM and ANN metrics.

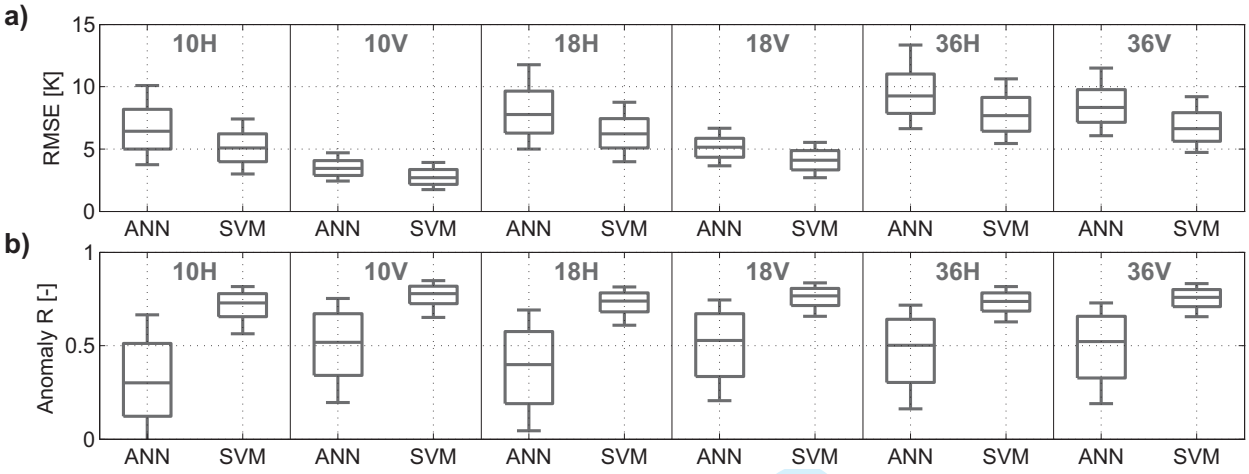

Fig. 3. Statistical box plots of a) $RMSE$ and b) anomaly R across the North American domain for the ANN and SVM from 1 September 2002 through 1 September 2011. Statistics are computed relative to AMSR-E measurements *not* used during training. Each box represents the median along with the 25th- and 75th-percentiles while the whiskers illustrate the 10th- and 90th-percentiles.

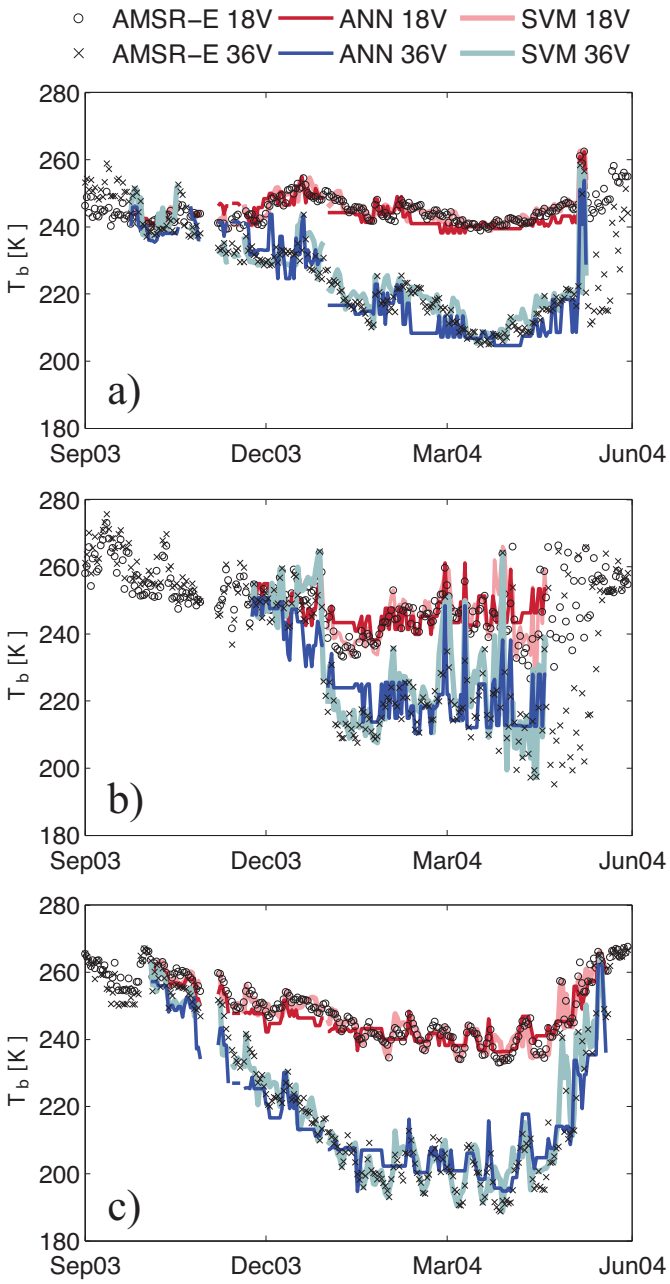

Fig. 4. Time series from 1 September 2003 through 1 Jun 2004 including AMSR-E observations, ANN estimates, and SVM estimates at 18V and 36V. a) Location with shallow snow depth and no forest cover (max. SWE = 0.07 cm; FF = 0.0; lat = 66.5°; lon = -66.7°). b) Location with moderate snow depth and high forest cover (max. SWE = 0.13 m; FF = 0.89; lat = 52.4°; lon = -85.1°). c) Location with large snow depth and modest forest fraction (max. SWE = 0.22 m; FF = 0.02; lat = 67.6°; lon = -151.6°). See also Figure 1 for locations.

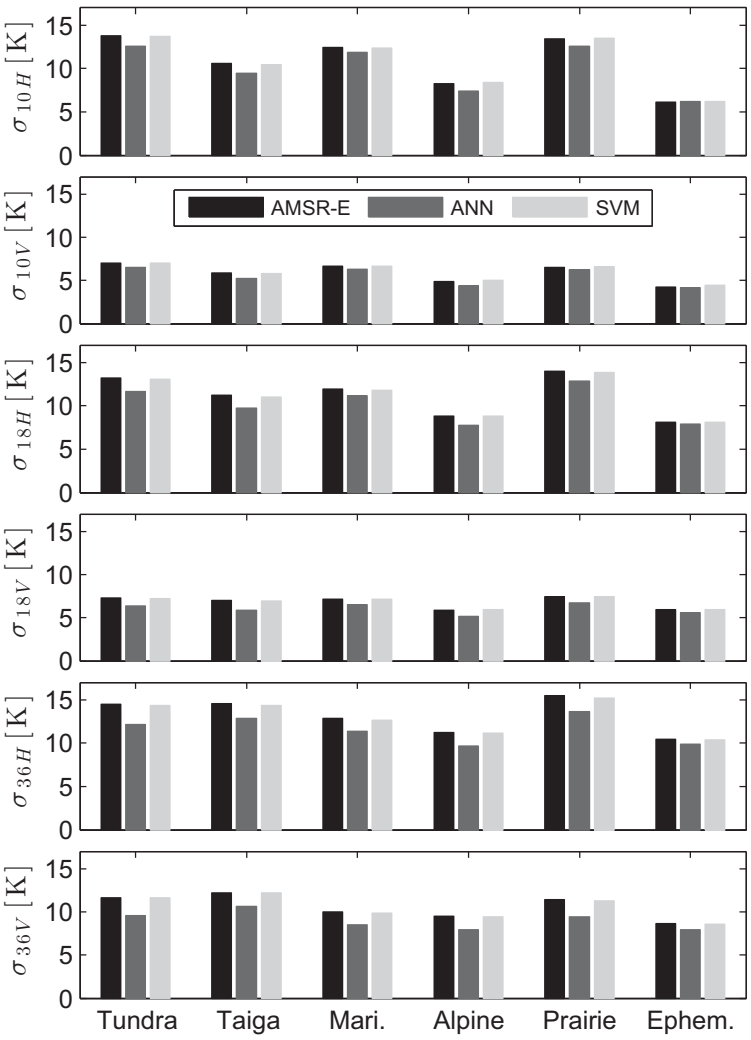

Fig. 5. Spatial variability (σ) of (black) AMSR-E, (dark gray) ANN, and (light gray) SVM T_bs time-averaged by snow class according to [33] for the 9-year study period for all frequency and polarization combinations.

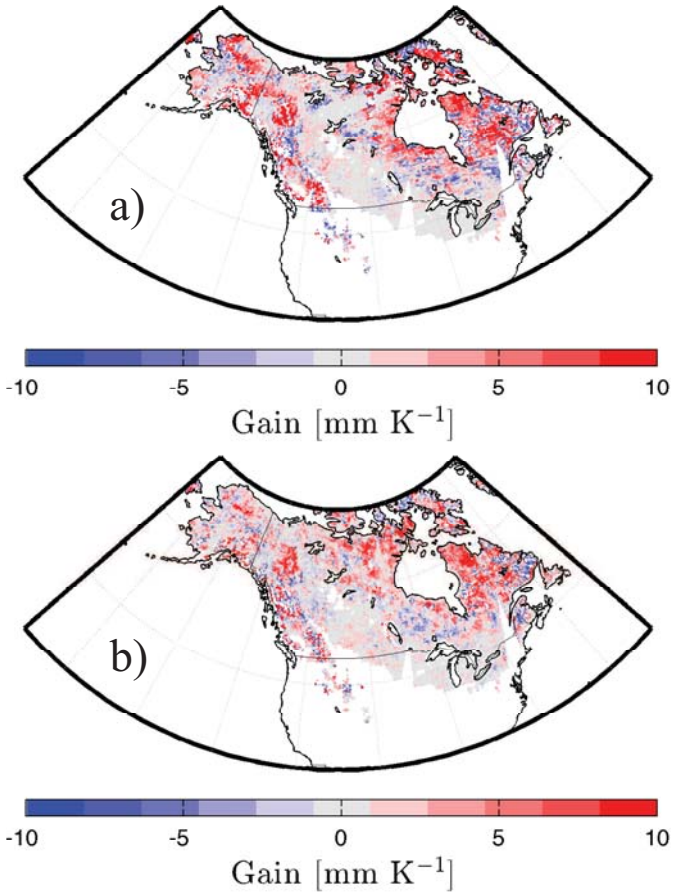

Fig. 6. Kalman gain for SWE versus ΔT_b=18V-36V near peak SWE accumulation on 1 February 2003 for a) ANN-derived estimates and b) SVM-derived estimates.

Barton A. Forman received his B.S. (cum laude) in Civil Engineering from the University of Virginia, his M.S. in Civil and Environmental Engineering from the University of California at Berkeley, and his Ph.D. in Civil Engineering from the University of California at Los Angeles where he received the Edward K. Rice Outstanding Ph.D. Student of the Year award. Upon completion of his graduate studies, he was awarded a NASA Postdoctoral Fellowship at the NASA Goddard Space Flight Center where his work focused on the estimation of snowpack using space-based measurements of the Earth's gravitational field. He is currently an Assistant Professor in the Department of Civil and Environmental Engineering at the University of Maryland.

Rolf H. Reichle received the M.S. degree ("Diplom") in Physics from the University of Heidelberg, Germany, in 1996 and the Ph.D. degree in Environmental Engineering from the Massachusetts Institute of Technology in 2000. He is currently a Research Physical Scientist with the NASA Global Modeling and Assimilation Office in Greenbelt, Maryland. His research interests include land data assimilation, satellite-based remote sensing, and applications related to land-atmosphere interactions, weather prediction and short-term climate forecasting.